J Stat Phys manuscript No.
(will be inserted by the editor)

Cyril Furtlehner · Jean-Marc Lasgouttes · Maxim Samsonov

One-dimensional Particle Processes with Acceleration/Braking Asymmetry

Received: date / Accepted: date

Abstract The slow-to-start mechanism is known to play an important role in the particular shape of the Fundamental Diagram of traffic and to be associated to hysteresis effects of traffic flow. We study this question in the context of exclusion and queueing processes, by including an asymmetry between deceleration and acceleration in the formulation of these processes. For exclusions processes, this corresponds to a multi-class process with transition asymmetry between different speed levels, while for queueing processes we consider non-reversible stochastic dependency of the service rate w.r.t. the number of clients. The relationship between these 2 families of models is analyzed on the ring geometry, along with their steady state properties. Spatial condensation phenomena and metastability are observed, depending on the level of the aforementioned asymmetry. In addition, we provide a large deviation formulation of the fundamental diagram which includes the level of fluctuations, in the canonical ensemble when the stationary state is expressed as a product form of such generalized queues.

1 Introduction

In the study of models for traffic [25], properties of the Fundamental Diagram (FD), which gives a relation between the traffic flow and the density or alternatively the dependence between the speed and the flow, or the speed and the density, play an important role. In the three phases traffic theory of Kerner [17], the traffic phase diagram on highways consists of the free flow, the synchronized flow and the wide moving jam. In the free flow regime, at low density, the flow is simply

C. Furtlehner and M. Samsonov
INRIA Saclay–Île-de-France – LRI, Bat. 490, Université Paris-Sud – 91405 Orsay cedex (France)

J.-M. Lasgouttes
INRIA Paris–Rocquencourt – Domaine de Voluceau B.P. 105 – 78153 Le Chesnay cedex (France)

proportional to the density of cars; in the congested one, at large density, massive clusters of cars are present, and the flow decreases more or less linearly with this density; in the intermediate regime, the relation between flow and density is largely of stochastic nature, due to the presence of a large amount of small clusters of cars propagating at various random speeds. It is not clear however whether in this picture these phases, and especially the synchronized flow phase, are genuine dynamical or thermodynamical phases, meaningful in some large size limit in the stationary regime, or are intricate transient features of a slowly relaxing system. In fact, there is still controversy about the reality of the synchronized phase of Kerner at the moment [24].

In order to analyze properties of the FD from a statistical physics perspective, we look for a simple extension of the totally asymmetric exclusion process (TASEP) as well as the zero range process (ZRP) able to encode the fact that vehicles in the traffic may accelerate or brake. Some asymmetry between these two actions is empirically known to be responsible for the way spontaneous congestion occurs. This is referred to as the slow-to-start mechanism, which is not explicitly present in the original cellular automaton of Nagel-Schreckenberg [21], but is part of its refined versions like the velocity dependant randomization (VDR) model [2]. VDR exhibits in particular a first order phase transition between the fluid and the congestion phase and some hysteresis phenomena [3] associated to metastable states. A realistic stochastic model should therefore have the property that some spontaneous symmetry breaking among identical vehicles may occur, as can be seen experimentally on a ring geometry for example [28]. With such a model at hand, we would like to provide a method to compute the FD and study emergence of non-trivial collective behaviors at macroscopic level.

The paper is organized as follows: in Section 2 we start by defining a simple and somewhat minimal generalization of TASEP and describe some of its basic properties. In Section 3 a family of ZRP is introduced, which service rate depends stochastically on the state of the queue, and on which there are relevant mappings of multi-speed TASEP; we determine sufficient conditions for having a product form for the invariant measure of such processes coupled in tandem. In Section 4, we provide a large deviation formulation giving the FD along with fluctuations on the ring geometry in the canonical ensemble, when the steady state has a product form. Finally, Section 5 is devoted to the analysis of the jam structure of a generalized queueing process corresponding to the model introduced in Section 2 and to the interpretation of direct simulations of the multi-speed TASEP in this light.

2 The Acceleration/braking-TASEP (AB-TASEP)

2.1 Model definition

The model we investigate is a multi-type exclusion process, generalizing the simple exclusion process on the line, introduced by Spitzer in the 70's [19,27], combined with some features of the Nagel-Schreckenberg cellular automaton [21].

In the totally asymmetric version of the exclusion process [19,27] (TASEP), particles move randomly on a 1-d lattice, always in the same direction, hopping from one site to the next one within a time interval following a Poisson distribution and conditionally that the next site is vacant. In the Nagel-Schreckenberg cellular

automaton, the dynamics is in parallel: all particles update their positions at fixed time intervals, but their speeds are encoded in the number of steps that they can take. This speed can adapt stochastically, depending on the available space in front of the particle.

The model that we propose combines the braking and accelerating feature of the Nagel-Schreckenberg models, with the locality of the simple ASEP model, in which only two consecutive sites do interact at a given time. The trick is to allow each car to change stochastically its hopping rate, depending on the state of the next site. For a 2-speed model, let A (resp. B) denote a site occupied by a fast (resp. slow) vehicle, and let O denote an empty site; the model is defined by the following set of reactions, involving pairs of neighbouring sites:

$$AO \xrightarrow{\mu_a} OA \qquad \text{simple move of fast vehicle} \tag{1}$$

$$BO \xrightarrow{\mu_b} OB \qquad \text{simple move of slow vehicle} \tag{2}$$

$$BO \xrightarrow{\gamma} AO \qquad \text{slow vehicle spontaneously accelerates} \tag{3}$$

$$A\bar{O} \xrightarrow{\delta} B\bar{O} \qquad A \text{ brakes behind } \bar{O} = A \text{ or } B \tag{4}$$

μ_a, μ_b, γ and δ denote the transition rates of the associated Markov process. The dynamics is therefore purely random sequential, as opposed to the parallel dynamics of the Nagel-Schreckenberg model. It encodes the tendency of a vehicle to accelerate when there is space ahead (3), and to slow down otherwise (4). The main mechanism behind congestion, namely the asymmetry between braking and acceleration is potentially present in the model when γ is different from δ. Our model is in fact similar to the model of Appert-Santen [1], in which there is a single speed, but particles have 2 states (at rest and moving), with possible transitions between these 2 states.

To define fully the model, its boundary conditions have to be specified, either periodic in the ring geometry or with edges. In the latter case, additional incoming and outgoing rates have to be specified, depending on whether we want to model a traffic light, a stop sign or simply a segment of highway for example.

In this paper, we will consider the model on the ring geometry with a total size denoted $S = N + L$, N being the (fixed) number of vehicles and L the number of empty sites. The rules (1)-(4) will be referred to as the acceleration-braking totally asymmetric exclusion process (AB-TASEP). This model contains and generalizes several sub-models which are known to be integrable with particular rates. The hopping part (1,2) of the models is just the TASEP when $\mu_a = \mu_b$, which is known to be integrable with help of the Bethe Ansatz (see e.g. [11] and reference therein). Its generalization to include multiparticle dynamics with overtaking is the so-called Karimipour model [4, 14], which turns out to be integrable as well. The Matrix Ansatz method allows one, in some cases like the ASEP with open boundary conditions, to describe the stationary regime of the model, using the representations of the so-called diffusion algebras [5]. The acceleration/deceleration dynamics is equivalent to the coagulation/decoagulation models, which are known to be solvable by the empty interval method and by free fermions for particular sets of rates [23], but the whole process is presumably not integrable.

2.2 Relation to tandem queues

We have shown in [10] that the model can be exactly reformulated in terms of a queueing network, where service rate of each server follows as well a stochastic dynamics. In this previous work, we however considered exclusion processes involving three consecutive sites interactions, in order to maintain indentical labels inside particle clusters. The mechanism is that, when a particle gets in close contact with the next one, it adopts its rate instantaneously, and when it leaves a cluster, it adopts a new rate at random. The mapping works only on the ring geometry, by identifying servers either with

(i) cars: clients are the empty sites,
(ii) empty sites: clients are the vehicles.

In the present case, the mapping of type (i) is exact up to a slight extension of the ordinary definition of a queueing process. In this new process by contrast to standard ones, the service rate is a stochastic Markov process as well: it can take two values $\mu_a > \mu_b$, and slow queues become fast at rate γ conditionally to having at least one client, while empty fast queues become slow at rate δ.

The mapping of type (ii) is more informative with respect to jam distribution, but is not possible with transitions limited to 2-consecutive sites interactions, because in that case homogeneity is not maintained inside clusters, car may have different labels and additional information besides the number of cars and the speed of the car leaving the queue is needed to know the service rate of the server. For this mapping, we have to resort to some approximate procedure that shall be described in Section 5.

Another way to circumvent this problem would be to start from a slightly different definition of our initial model, by not attaching speed labels to cars, but instead to empty sites. Let us just give an example of such rules. If F and S denote now empty site with respectively fast and slow label speed, and V denote sites occupied with a vehicle, the set of transitions reads then

$$\begin{aligned} VF &\xrightarrow{\mu_f} FV \quad \text{simple move on fast site} \\ VS &\xrightarrow{\mu_s} SV \quad \text{simple move on slow site} \\ VF &\xrightarrow{\delta} VS \quad \text{fast site become slow} \\ \bar{V}S &\xrightarrow{\gamma} \bar{V}F \quad \text{slow site become fast} \end{aligned} \tag{5}$$

where $\bar{V}$ is an unoccupied site i.e. either F or S. The mechanism in this model is that the empty sites visited more recently are associated to a slower speed than others, leading possibly to congestion instability. In fact, these rules lead to a model very similar to the Bus Route Model (BRM) studied in [22]. The only difference is that the acceleration/braking rule is reversed, in the sense that sites become fast when not visited, in contrary to what occurs in the BRM.

The mapping to tandem queues is then straightforward: queues are associated to empty sites of type F or S with corresponding service rate μ_f or μ_s; fast queues with at least one client have a probability per unit of time δ to become a slow

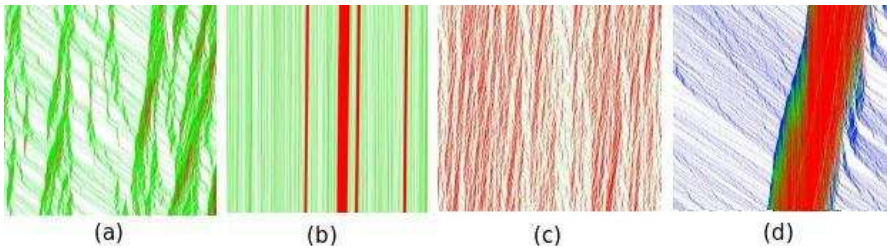


Fig. 1 (Color online) (a)-(c) Space-time diagram of AB-TASEP. Time is going downward and particles travel to the right. Slow ones ($\mu_b = 10$) are colored in red and fast ones $\mu_a = 100$ in green. The system size is $S = 3000$ for (a) and (c) and $S = 100000$ for (b), all with same density $\rho = 0.2$. Other parameters are $\gamma = 10$, $\delta = 1$ for (a) and (b) and $\delta = 10$ for (c). (d) is a jam obtained for a model with an additional label C colored in blue, with additional transition rules similar to (2,3,4) corresponding to a faster speed level μ_c.

queue, while slow queue which are empty become fast with probability rate γ. This model generalizes directly to any number of speed levels. Interestingly, one sees that, in such a model, the service rate of queues are somehow related to the age of clusters. Note also that the BRM can as well be exactly mapped to such a process.

To avoid confusions between various models, we focus in this paper on the study of the AB-TASEP model of Section 2.1, postponing to future work the study of model (5).

2.3 Numerical observations

Based on numerical simulations on the ring geometry, we make some observations concerning the phenomenology of the model (1)-(4), depending on the parameters. This is illustrated on Figure 1. When no asymmetry between braking and accelerating is present ($\gamma = \delta$), as in TASEP on a ring, no spontaneous large jam structure is observed. As the density $\rho \stackrel{\text{def}}{=} N/S$ of cars increases, one observes a smooth transition between a TASEP of fast particles for small ρ to a TASEP of slow particles around $\rho \simeq 1$. Instead, when the ratio δ/γ is reduced, there is a proliferation of small jams. Below some threshold of this ratio, we observe a condensation phenomenon, associated to some critical value of the density: above this critical value, after a time which presumably scales as some power of S, one or more large jams absorbing a finite fraction of the total number of cars may develop. If there are many of such large jams, a competition occurs, which ends up possibly in the long term with one single large condensate depending on the relative value of γ and μ_b and also of the size S of the system.

It is tempting to interpret this as a condensation mechanism at equilibrium in the canonical ensemble [7], by combining the Nagel-Paczuski [20] interpretation of competing queues with some results [15, 10] which, in the context of tandem queues on a ring, allows this condensation mechanism to take place if the appearance of slow vehicle is a sufficiently rare event. We will come back at depth in Section 5 to the queueing interpretation of jam formation in the AB-TASEP model. For the moment, let us analyze this qualitatively.

Consider a jam of size n and let t_n be the expected waiting time in this queue. Assuming by convention that particles move to the right, t_n is by definition the time it will take until the left end particle performs its next move. It corresponds to the sum of n exponential events, where each particle leaves the jam in turn. The time it takes for each individual event depends on the label of the particle at the moment where it becomes the leading particle in the queue. For a particle of type A entering the queue of size n, the probability that it remains of this type when reaching the foremost position in the queue is simply

$$p(\mu = \mu_a) = e^{-\delta t_n} = p_n.$$

If we assume that n does not change much during t_n, then each particle that reaches the front place has been waiting for a time t_n, and therefore

- is still of type A with probability p_n and leaves the queue within a mean time $1/\mu_a$;
- is of type B with probability $1-p_n$ and leaves the queue within a mean time $1/(\mu_b+\gamma)+\gamma/(\mu_b+\gamma)\times 1/\mu_a$, if we take into account the fact that the particle may accelerate at rate γ before leaving the queue.

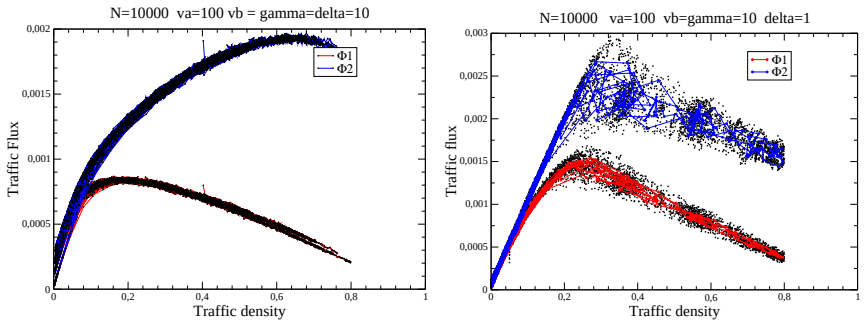
We get therefore the rough estimate

$$t_n \simeq n\Big(\frac{1}{\mu_a}p_n + \frac{\mu_a+\gamma}{\mu_a(\mu_b+\gamma)}(1-p_n)\Big),$$

giving t_n self-consistently. This indicates qualitatively that, if at the beginning of the process, there is a distribution of jams with various effective service rates, as time evolves, long-lived jams with effective service rates close to $\mu(n\to\infty) = \mu_a(\mu_b+\gamma)/(\mu_a+\gamma)$ are more likely to survive; this gives an upper bound to the intensity of incoming particles into the other jams because we are on the ring geometry. Hence, a new small jam cannot increase in size very much because its effective service rate is strictly larger than this limit, so there is coexistence of very small jams with the existing large jams; the largest one eventually remains alone after waiting for a large amount of time. This is typically observed when $\delta \ll \mu_b$ and $\delta < \gamma$ (Figure 1.b). If the situation with a single jam is not stable, then no large jam may develop at all, and only small fluctuations are to be observed. When $\delta > \gamma$ one likely observes the kind of jams seen in Figure 1.c.

In [13] a general criterion for having phase separation is conjectured for conserved systems, based on the asymptotic behavior of the current passing through clusters of large size. Given t_n above, we expect an exponential decrease to the limit current $J_\infty = \mu(n\to\infty)$, much faster than the limit power law behavior, $J_n \simeq J_\infty(1+b/n)$, $b<2$, below which no condensation should occur for infinite system size, even at high density. Thus, according to this criterion, the observed condensate should be a finite size effect.

A way to observe the effect of asymmetry is to allow particles to enter or quit the system with some very low rate when compared to the others. The global density of cars then performs a random walk, and by looking at trajectories in the FD plane, we see hysteresis effect for $\delta < \gamma$. Two different quantities are interesting



(a) (b)

Fig. 2 (Color online) Fundamental diagram with $\gamma = \delta$ (a) and $\gamma = 10\ \delta$ (b) with small spontaneous rate of emission and escape of particles. Trajectories of Φ_1 and Φ_2 are displayed. Φ_1 is the actual particle flow and Φ_2 is interpreted as the "traffic flow".

here:

$$\Phi_1 \stackrel{\text{def}}{=} \frac{1}{S}\sum_{i=0}^{S-1} \mu_a A_i O_{i+1} + \mu_b B_i O_{i+1},$$

$$\Phi_2 \stackrel{\text{def}}{=} \frac{1}{S}\sum_{i=0}^{S-1} \mu_a A_i + \mu_b B_i,$$

(with Boolean variables $A_i + B_i + O_i = 1$). Φ_1 represents the flow of particles, while Φ_2 counts particles with their speed, regardless of whether the next site is occupied or not. In a way, Φ_2 is more representative of the traffic flow than Φ_1 because in practice cars follow a parallel dynamics and cluster of particles should represent moving platoons of cars. Indeed, this is reflected in the FD of Figure 2.b, where Φ_2 gives a much more realistic FD than Φ_1, which produces the typical and non-realistic round shape due to the sequential dynamics. This suggests that, when comparing the FD diagram of sequential exclusion processes with parallel dynamical cellular automata, Φ_2 should be considered rather than Φ_1.

We have also simulated a model with 3 speed levels, shown on Figure 1.d. In this case, small jams may have different speeds, depending on which type of slow car is leading. Then a cascade mechanism takes place, slow speed regions generate even slower speed clusters of cars and so on, and some kind of synchronized flow is observed.

To conclude this section, let us finally remark that Figure 1.a is very reminiscent of a coagulation-decoagulation process, which is somewhat expected from the previous discussion in Section 2.1, since it is present in the equations. One could therefore expect the jam structure to share some properties with the directed percolation process taking place on the ring geometry.

2.4 Hydrodynamic equations and solitons

The hydrodynamic limit of TASEP has been established under rather general conditions [18], and can be extended to situations with several conservation laws [29]. In the present case, there are two species but only one conservation law. While, it is not clear whether the hydrodynamic limit remains valid in this context, let us write down the expected equations in such a case. They deal with the density

$$(\rho_a(x,t),\rho_b(x,t)) = \lim_{S\to\infty} \mathbb{E}\big(A^{St}_{[xS]}, B^{St}_{[xS]}\big)$$

obtained by a rescaling on both the space and time of the pair of process $(A^t,B^t) = \{(A^t_i,B^t_i), i = 0,\dots,S-1\}$ where $A^t_i \in \{0,1\}$ and $B^t_i \in \{0,1\}$ represents respectively the presence of a particle of type A ($A^t_i = 1$) and of type B ($B^t_i = 1$) on site i at time t, and the exclusion constraint reads $A^t_i + B^t_i \in \{0,1\}$. Writing down the conservation equations for $\mathbb{E}\big(A^t_i\big)$ and $\mathbb{E}\big(B^t_i\big)$ and neglecting the correlations between neighboring sites yields the following coupled system when $S \to \infty$:

$$\partial_t\rho_a + v_a\partial_x(\rho_a(1-\rho)) = -\delta\rho_a\rho + \gamma\rho_b(1-\rho), \tag{6}$$

$$\partial_t\rho_b + v_b\partial_x(\rho_b(1-\rho)) = \delta\rho_a\rho - \gamma\rho_b(1-\rho), \tag{7}$$

with $\rho = \rho_a + \rho_b$ the total density of cars and

$$v_a \overset{\text{def}}{=} \lim_{S\to\infty} \frac{\mu^{(S)}_a}{S}, \qquad v_b \overset{\text{def}}{=} \lim_{S\to\infty} \frac{\mu^{(S)}_b}{S}$$

since it makes sense to have rates $\mu^{(S)}_a$ and $\mu^{(S)}_b$ linearly growing with S in hydrodynamic limit. This is an hyperbolic system of equations since the matrix

$$M = \begin{bmatrix} v_a(1-2\rho_a-\rho_b) & -v_a\rho_a \\ -v_b\rho_b & v_b(1-\rho_a-2\rho_b) \end{bmatrix}$$

has two real eigenvalues

$$\lambda^{\pm} = \frac{1}{2}\big(v_a + v_b - v_a(2\rho_a+\rho_b) - v_b(\rho_a+2\rho_b) \pm \sqrt{\Delta},$$

with

$$\Delta = \big(v_a(1-2\rho_a-\rho_b) - v_b(1-\rho_a-2\rho_b)\big)^2 + 4v_av_b\rho_a\rho_b \ > 0.$$

The method of characteristics applied to this system basically amounts to find the family of characteristic curves $x(t,x_0)$, indexed by $x_0 \in [0,1]$, given e.g. an initial density profile $\rho_0(x)$ and along which ρ remains constant. It is however not giving much insight, because here these curves have no reason to be straight lines. Instead, the expression of $\lambda^{\pm}$ shows that they actually depend on the second degree of freedom $\rho_a - \rho_b$, which is not conserved by the dynamics and which cannot be eliminated easily from the equations.

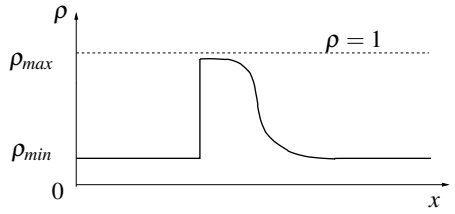


Fig. 3 Density profile of a solitary wave solution to equations (6,7).

Nevertheless, we can look at steady state for backward travelling waves solutions of the form $\rho_{a,b}(x+vt)$, with v the propagation velocity. Replacing ∂_t by $v\partial_x$ in (6,7) and summing up the two equations, one can check that the flow

$$\Phi \stackrel{\text{def}}{=} (v_a\rho_a + v_b\rho_b)(1-\rho) + v\rho, \tag{8}$$

is a conserved quantity

$$\frac{d\Phi}{dx} = 0.$$

This allows one to express ρ_a and ρ_b as function of ρ:

$$\rho_a = \frac{\Phi - \rho v - v_b\rho(1-\rho)}{(v_a - v_b)(1-\rho)} \qquad \text{and} \qquad \rho_b = \frac{\Phi - \rho v - v_a\rho(1-\rho)}{(v_b - v_a)(1-\rho)} \tag{9}$$

After rescaling w.r.t. v_a, the model has only 3 independent parameters, like e.g.

$$\eta \stackrel{\text{def}}{=} \frac{v_b}{v_a}, \qquad \kappa \stackrel{\text{def}}{=} \frac{\delta}{\gamma}, \qquad \nu \stackrel{\text{def}}{=} \frac{\gamma}{v_b},$$

and we rescale as well the soliton speed $\omega \stackrel{\text{def}}{=} v/v_a$ and the flow $\phi \stackrel{\text{def}}{=} \Phi/v_a$. Replacing ρ_a and ρ_b with their expressions in (9), we obtain an equation relating $\partial_x\rho$ and ρ:

$$\frac{d\rho}{dx} = -C(1-\rho)\,\frac{P(\rho)}{Q(\rho)},$$

with

$$C = \frac{\nu}{2}(1-\eta\kappa)$$

and where P and Q are both polynomials of degree 3, which can be written as:

$$P(\rho) \stackrel{\text{def}}{=} (1-\rho)(\rho-\rho^-)(\rho-\rho^+) + \big(\frac{\nu}{2}(1-\kappa)(1-\rho) + \kappa\nu\big)(\phi-\omega)$$

$$Q(\rho) \stackrel{\text{def}}{=} (1-\rho)^2(\rho-\rho_1) + (\phi-\omega)\frac{\omega}{2\eta}$$

where the roots of P and Q when $\phi = \omega$ are

$$\rho_1 = 1 + \frac{1+\eta}{\eta}\omega, \qquad \rho^\pm = \frac{1 + (1-\kappa)\omega \pm \sqrt{\Delta}}{2(1-\eta\kappa)},$$

with the discriminant, assumed to be positive,

$$\Delta = (1-(1+\kappa)\omega)^2 + 4(\eta-\omega)\kappa\omega.$$

Considering $\kappa < 1$, we have $\rho^- \in [0,1]$ and $\rho^+ \geq 1$ when $\omega \leq \eta$ and $\rho^+ \in]\rho^-, 1]$ when $\omega \geq \eta$. $1/C$ gives the length scale of the interface between regions of different densities. In particular, if μ_a, μ_b and γ are maintained at fixed values, the interface becomes a step function when S is large since $v_b \to 0$.

Let $\rho_{\max}$ and $\rho_{\min}$ be the extremal values of the density. We expect a solution which decreases from $\rho_{\max}$ to $\rho_{\min}$ with $R(\rho) \stackrel{\text{def}}{=} P(\rho)/Q(\rho)$ staying strictly positive for $\rho \in [\rho_{\min}, \rho_{\max}]$. Since $R(\rho)$ is a single valued function of ρ, there must be a discontinuity, i.e. a shock to relate these two values as depicted on Figure 3. We do not expect everything to be discontinuous at this point, in particular, as long as the shock is not at rest ($v = 0$) the proportion of fast cars $u \stackrel{\text{def}}{=} \rho_a/\rho$ should be continuous, since cars are slowing down at finite pace δ. This means that $\rho_{\min}$ and $\rho_{\max}$ are not independent. Instead, letting u^* be the value of u at the shock, from (9), they are both solution of

$$(v_a u^* + v_b(1-u^*))\rho^2 - (v_a u^* + v_b(1-u^*) + \omega)\rho + \phi = 0.$$

Given that the sum and product of the two roots have to be smaller than 2 and 1 respectively, we deduce that ω and ϕ are both strictly smaller than $v_a u^* + v_b(1-u^*)$. In addition the constraint $\rho_{\max} \leq 1$ implies $\phi \geq \omega$. As a consequence, for $\rho \in [0,1[$ we have $Q(\rho) > 0$, vanishing only at $\rho = 1$ in the limit case $\Phi = \omega$. Concerning $P(\rho)$, it has then 3 roots $\rho_i, i \in \{1,2,3\}$ with $0 < \rho_1 < \rho_2 < 1 < \rho_3$. To obtain a general solution, there are two free parameters, ω and $\rho_{\min}$ and therefore 2 constraints, associated to fixing the length of the system to 1 and the mean density to ρ_0. These read

$$\int_{\rho_{\min}}^{\rho_{\max}} \frac{d\rho}{(1-\rho)} \frac{Q(\rho)}{P(\rho)} = C \qquad \text{and} \qquad \int_{\rho_{\min}}^{\rho_{\max}} d\rho \frac{\rho}{(1-\rho)} \frac{Q(\rho)}{P(\rho)} = C\rho_0.$$

When condensation occurs, we expect to observe a full congestion ($\rho = 1$) at some point. This entails from (8) that $\phi = \omega$, yielding in that case the equation:

$$\frac{d\rho}{dx} = C\frac{(\rho^+ - \rho)(\rho - \rho^-)}{\rho - \rho_1}.$$

and

$$\rho_a = \frac{\omega - \eta\rho}{1-\eta} \qquad \text{for} \qquad \rho < 1.$$

On one hand, we should have $\rho^+ > 1$ to be able to have a point where $\rho = 1$, consequently ω is necessarily smaller than η; on the other hand ρ_a becomes negative when $\rho \to 1$ if $\omega < \eta$. As a result, genuine condensation which signature would be the presence of a shock with $\rho_{\max} = 1$ is not obtained in this hydrodynamic framework. At this point, it not clear how to interpret this fact:

- either the hydrodynamic equations are simply not correct because of non-vanishing local correlations when $S \to \infty$;

– or the hydrodynamic limit ceases to be valid in the condensation regime, in contrary to what happens for example for the ordinary ZRP, where the hydrodynamic limit can be extended in this case [26].
– or the condensation which is observed numerically is a finite size artifact which does not survive in the limit $S \to \infty$.

As we shall see, the forthcoming analysis of the model based on queueing process will be in favour of this last interpretation.

3 Tandem queues with stochastic service rates

3.1 Definition of the process

As already mentioned in Section 2.2, in some cases on the ring geometry, exclusion processes can be mapped onto generalized queueing processes which we define in the present section. They have a fixed number of servers, organized in tandem and in each queue, both the number of clients and the service rate are stochastic, the service rate being not necessarily a function of the number clients. Taken in isolation, each queue may not be reversible, but instead presents some hysteresis feature which might be relevant to traffic. It is a generalization of queueing processes with stochastic service [12, 10, 23]. From statistical physics viewpoint, it is an extension to the ZRP, obtained by adding internal dynamics.

This family of models is defined more formally as follows: let $\mathscr{G} = (\mathscr{N}, \mathscr{L})$ a network of queues with dynamical stochastic service rates. The state of each server $i \in \mathscr{N}$ is represented by a state vector $z_i(t) = (n_i(t), \mu_i(t)) \in E_i \subset \mathbb{N} \times \mathbb{R}^+$, where $n_i(t)$ is the number of clients and $\mu_i(t)$ is a service rate. Let $V_i^-(z_i)$, $V_i^+(z_i)$ and $V_i^0(z_i)$ be the sets of states in E_i having respectively $n_i - 1$, $n_i + 1$ and n_i clients. Considering server i alone, any transition from state z_i ends up necessarily in one of these sets; for example, the service rate $\mu_i(t)$ represents the transition rate from z_i to $V_i^-(z_i)$. Two sets of transition probability matrices $p_i^{\pm}(z,z')$ and one set of transition rates $q_i^0(z,z')$ have to be introduced to complete the definition of the individual queueing process. When a client gets served in queue i, there is a transition from the state of the departure queue z_i to one of the states $z' \in V_i^-(z)$ with a rate $\mu_i(t) p_i^-(z,z')$; in the destination queue, a transition occurs from the state z_{i+1} into one of the states $z' \in V_{i+1}^+(z)$ with probability $p_{i+1}^+(z,z')$. We have the normalizations,

$$\sum_{z' \in V_i^+(z)} p_i^+(z,z') = \sum_{z' \in V_i^-(z)} p_i^-(z,z') = 1. \tag{10}$$

Additional internal transitions are allowed, in which the service rate μ_i of server i changes independently of any arrival or departure. The intensities of these transitions are given by the set $\{q_i^0(z,z'),\ z' \in V_i^0(z)\}$ of transition rates. To summarize, the combined set of transition rates reading

$$q_i(z,z') \stackrel{\text{def}}{=} \lambda\, p_i^+(z,z') \mathbb{1}_{\{z' \in V_i^+(z)\}} + \mu(z) p_i^-(z,z') \mathbb{1}_{\{z' \in V_i^-(z)\}} + q_i^0(z,z') \mathbb{1}_{\{z' \in V_i^0(z)\}},$$

defines for each $i \in \mathscr{N}$, a continuous time Markov process representing the dynamics of one queue taken in isolation with arrival rate λ. When the servers are

arranged into a tandem of queues, the output of server i becomes the input of server $(i+1) \mod (L)$ with L the number of servers. For more general arrangements, when all the servers are identical, the joint process is entirely specified by:

- the network of queues $\mathscr{G}$;
- the routing matrix containing the probability weights of the various destinations from each server given by $\mathscr{G}$;
- the state graph of the single queue, which set of nodes $\{(n,\mu)\}$ is a subset of $\mathbb{N}\times\mathbb{R}^+$, and in which each transition is represented by an oriented edge (see Figure 5.a and 5.b).

3.2 Product form of clusters at steady-state

In the stationary regime, the invariant measure of a network of reversible queues is known to factorize into a product form under general conditions [16], which allows one to compute many equilibrium quantities explicitly. With a dynamical rate, it is interesting to check for such properties and in this section we study the conditions under which the stationary state has a product form. For simplicity, we restrict the analysis to the case where $\mathscr{G}$ corresponds to a (closed) tandem of L queues, where the clients leave queue i to enter in queue $(i+1) \mod (L)$, for any $i = 0,\ldots,L-1$. We can establish the following sufficient conditions:

Theorem 1 *Let π_i^λ denote the steady state probability corresponding to queue i taken in isolation and fed with a Poisson process with rate λ. If the following partial balance equations are satisfied (see Figure 4),*

$$\sum_{z\in V_i^+(z_i)} \mu(z)p_i^-(z,z_i)\pi_i^\lambda(z) = \lambda\pi_i^\lambda(z_i), \tag{11}$$

$$\sum_{z\in V_i^-(z_i)} \lambda p_i^+(z,z_i)\pi_i^\lambda(z) + \sum_{z\in V_i^0(z_i)} q_i^0(z,z_i)\pi_i^\lambda(z) = \Big(\mu(z_i) + \sum_{z\in V_i^0(z_i)} q_i^0(z_i,z)\Big)\pi_i^\lambda(z_i), \tag{12}$$

then the joint probability measure of state $\mathscr{S} \stackrel{\text{def}}{=} \{z_i, i\in\mathscr{N}\}$ has the following product form at steady state:

$$P\big(\mathscr{S} = \{z_i, i\in\mathscr{N}\}\big) = \frac{\prod_{i\in\mathscr{N}}\pi_i^\lambda(z_i)}{P(\sum_{i\in\mathscr{N}} n_i = N)}. \tag{13}$$

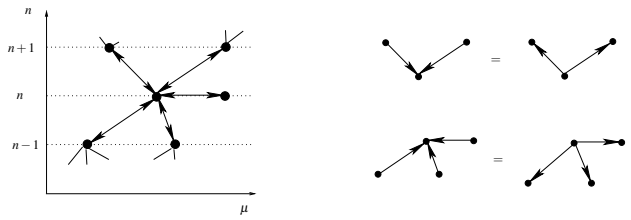


Fig. 4 Partial balance condition corresponding to (11) and (12).

Proof Using (10), the global balance equations reads

$$\sum_{i\in\mathscr{N}}\Bigg[\sum_{\substack{z'\in V_i^+(z_i)\\ z''\in V_{i+1}^-(z_{i+1})}}\mu(z')p_i^-(z',z_i)p_{i+1}^+(z'',z_{i+1})\frac{\pi_i^\lambda(z')\pi_{i+1}^\lambda(z'')}{\pi_i^\lambda(z_i)\pi_{i+1}^\lambda(z_{i+1})}$$

$$+\sum_{z\in V_i^0(z_i)}q_i^0(z,z_i)\frac{\pi_i^\lambda(z)}{\pi_i^\lambda(z_i)}\Bigg]P(\mathscr{S}=\{z_i\})$$

$$=\sum_{i\in\mathscr{N}}\Big[\mu(z_i)+\sum_{z\in V_i^0(z_i)}q_i^0(z_i,z)\Big]P(\mathscr{S}=\{z_i\}).$$

To find a sufficient condition, select term i in the first part of the left hand side of the above equation and terms $i+1$ in the remaining parts. This gives immediately, $\forall i\in\mathscr{N}$,

$$\sum_{\substack{z'\in V_i^+(z_i)\\ z''\in V_{i+1}^-(z_{i+1})}}\mu(z')p_i^-(z',z_i)p_{i+1}^+(z'',z_{i+1})\pi_i^\lambda(z')\pi_{i+1}^\lambda(z'')+$$

$$\sum_{z\in V_{i+1}^0(z_{i+1})}q_{i+1}^0(z,z_{i+1})\pi_i^\lambda(z_i)\pi_{i+1}^\lambda(z)$$

$$=\mu(z_{i+1})\pi_i^\lambda(z_i)\pi_{i+1}^\lambda(z_{i+1})+\sum_{z\in V_{i+1}^0(z_{i+1})}q_{i+1}^0(z_{i+1},z)\pi_i^\lambda(z_i)\pi_{i+1}^\lambda(z_{i+1}).$$

After multiplying this last equation by λ, this sufficient condition rewrites

$$\Big(\sum_{z'\in V_i^+(z_i)}\mu(z')p_i^-(z',z_i)\pi_i^\lambda(z')\Big)\Big(\lambda\sum_{z''\in V_{i+1}^-(z_{i+1})}p_{i+1}^+(z'',z_{i+1})\pi_{i+1}^\lambda(z'')\Big)$$

$$=\lambda\,\pi_i^\lambda(z_i)\times$$

$$\Big(\mu(z_{i+1})\pi_{i+1}^\lambda(z_{i+1})+\sum_{z\in V_{i+1}^0(z_{i+1})}q_{i+1}^0(z_{i+1},z)\pi_{i+1}^\lambda(z_{i+1})-q_{i+1}^0(z,z_{i+1})\pi_{i+1}^\lambda(z)\Big),$$

which is satisfied $\forall i\in\mathscr{N}$ under the two sets of sufficient conditions (11) and (12).

Note that these partial balance conditions (11,12) are only sufficient conditions for π_i^λ to be the invariant measure. They are however easier to check, because the proper parameters can be found directly from these equations. Note also that reversible queues are special cases of these rules and in this respect, this result is an extension to our context of the general results of Kelly concerning product forms in queueing networks [16].

The next obvious question is whether there exists non-reversible queues that satisfy conditions (11,12).

3.3 Examples

Let us give two examples of non-reversible isolated queues fulfilling these conditions leading to a product form.

Example 1: The state-graph corresponding to the first one is sketched on Figure 5.a. We use the following labelling of the states: $(z_0) = (0,0)$, $(z_1) = (1,\mu_1)$, $(z_2) = (1,\mu_2)$, $(z_3) = (2,\mu_4)$ and $(z_i) = (n-2,\mu_i)$ for $i \geq 4$. For $i \geq 4$, detailed balance is trivially verified. Imposing the partial balance conditions to the nodes $z_0,\ldots,z_3$ leads to four independent relations out of six, between the steady state probabilities π_i^λ of the isolated queue:

$$\lambda\pi_0^\lambda = \mu_1\pi_1^\lambda + \mu_2\pi_2^\lambda, \qquad \lambda\pi_1^\lambda = \mu_3 p_{31}\pi_3^\lambda,$$
$$q_{21}\pi_2^\lambda = (q_{12}+\mu_1)\pi_1^\lambda, \qquad \lambda\pi_2^\lambda = \mu_3 p_{32}\pi_3^\lambda,$$

which are actually compatible iff:

$$\frac{p_{32}}{p_{31}} = \frac{q_{12}+\mu_1}{q_{21}}.$$

Letting $p \stackrel{\text{def}}{=} p_{31} = 1 - p_{32}$, this corresponds to

$$p = \frac{q_{21}}{\mu_1 + q_{12} + q_{21}}.$$

When this condition holds, the joint measure of the tandem queues takes the product form (13), although each queue is by construction never reversible (transition $0 \to 1$ is absent). The invariant measure of the isolated queue reads

$$\pi_1 = \frac{\lambda q_{12}}{\mu_1\mu_2 + q_{12}\mu_2 + q_{21}\mu_1}\pi_0, \qquad \pi_2 = \frac{\lambda(\mu_1+q_{21})}{\mu_1\mu_2 + q_{12}\mu_2 + q_{21}\mu_1}\pi_0,$$
$$\pi_3 = \frac{\lambda^2(\mu_1+q_{12}+q_{21})}{\mu_3(\mu_1\mu_2 + q_{12}\mu_2 + q_{21}\mu_1)}\pi_0, \qquad \pi_{n\geq 4} = \left[\prod_{i=4}^{n}\frac{\lambda}{\mu_i}\right]\frac{\lambda q_{12}}{\mu_1\mu_2 + q_{12}\mu_2 + q_{21}\mu_1}\pi_0,$$

and π_0 is fixed by the normalization.

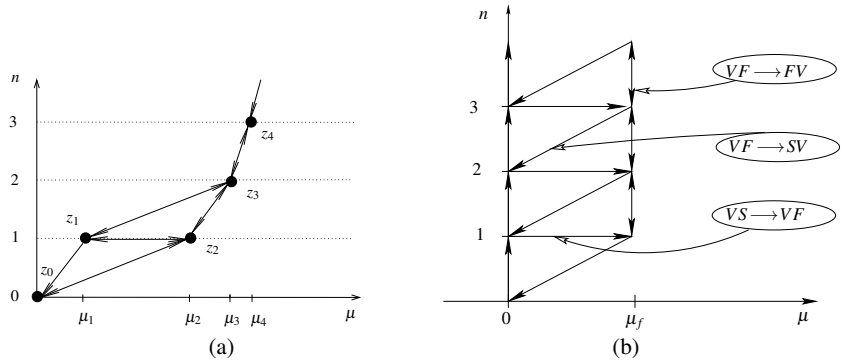


Fig. 5 Examples of a non-reversible single generalized queueing process obeying the partial balance condition (a) with exclusion process interpretation (b).

Example 2: Our second example is depicted on Figure 5.b. It is obtained from an exclusion process on a ring, which we define here for illustration purpose, even though it is not connected to AB-TASEP. Empty sites are of two different types, either blocking (S) or non-blocking type (F). Therefore, the particles (V) hop to the next site by following a two stage cascading process and the transition rules read:

$VF \xrightarrow{p\mu_f} FV$	vehicle hopping
$VF \xrightarrow{(1-p)\mu_f} SV$	vehicle hopping with site becoming blocking
$VS \xrightarrow{\mu_s} VF$	blocking site becomes non-blocking,

where $p \in [0,1]$ is a probability coefficient and μ_s and μ_f are transition rates. The corresponding queueing process is clearly non-reversible. Solving the partial balance equations leads to

$$\pi(n,\mu=0) = \rho_s^n \, \pi_0,$$

$$\pi(n,\mu=\mu_f) = \rho_f \frac{\rho_f^n - \rho_s^n}{\rho_f - \rho_s} \, \pi_0,$$

$$\pi_0 = (1-\rho_s)(1-\rho_f),$$

with $\rho_{s,f} = \lambda/\mu_{s,f}$ and an additional condition on the coefficient p:

$$\lambda \pi(n,\mu=\mu_f) = p\mu_f \pi(n+1,\mu=\mu_f).$$

This means that p actually depends on n like

$$p_{n+1} = \frac{1-(\mu_f/\mu_s)^n}{1-(\mu_f/\mu_s)^{n+1}} \in [0,1].$$

Strictly speaking, the original definition of Example 2 with a constant p does not lead to a zero range process with an exact product form, and we have not yet found any such example. Similarly, AB-TASEP, the Bus Route Model [22] and the model (5) discussed in Section 2.2 fail to meet our partial balance conditions. Such a product form can nevertheless be considered as a first approximation of mean-field type. The analysis of the fundamental diagram follows almost straightforwardly when such product form is present, except if the spectral gap between the steady-state and the first excited state vanishes. In that case, the fluctuations can be dominated by kinematic waves corresponding to the first excited states.

4 Fundamental diagram for product form distributions

4.1 Large deviation formulation

In practice, points plotted in experimental FD studies are obtained by averaging data from static loop detectors over a few minutes (see e.g. [17]). This is difficult to compute from our queue-based model, for which a space average is much easier to obtain. The equivalence between time and space averaging is not an obvious assumption [3], but since jams are moving, space and time correlations are combined in some way [20] and we consider this assumption to be quite safe. In this section, we want to extend the traditional study of the FD to the analysis of the fluctuations, i.e. the departure from the deterministic function relating the flow to the density. Experimentally, the congestion region of the FD is seen to be dominated by fluctuations, while the free flow part is rather deterministic. In what follows, we will therefore consider a probabilistic version of the FD, where the deviation from the deterministic FD i.e. the most probable point, will be analyzed in the large deviation framework. A large deviation speed corresponding to the number L of servers, instead of time t as is usually done for time-averaged flows, will be considered. The probabilistic measure which we consider is the conditional probability $P(\phi|d)$, where d represents the spatial density of cars and ϕ the normalized flow:

$$\begin{cases} d = \dfrac{N}{N+L}, \\ \phi = \dfrac{\Phi}{N+L}, \end{cases} \quad \text{with} \quad \begin{cases} L & \text{number of queues} \\ N = \displaystyle\sum_{i=1}^{L} n_i & \text{number of vehicles} \\ \Phi = \displaystyle\sum_{i=1}^{L} \mu_i \mathbb{1}_{\{n_i>0\}} & \text{integrated flow} \end{cases}$$

The numbers N of vehicles and L of queues are fixed, meaning in the statistical physics parlance, that we are working with the canonical ensemble. If we assume that we are in the conditions of having a product form for the stationary distribution, with individual probabilities $\pi^{\lambda}(n,\mu)$ associated to each queue taken in isolation, then, taking into account the constraints yields the following form of the joint probability measure:

$$P(\{n_i,\mu_i\}) = \frac{\delta\left(N-\sum_{i=1}^{L} n_i\right)}{Z_L(N)} \prod_{i=1}^{L} \pi^{\lambda}(n_i,\mu_i),$$

with canonical partition function

$$Z_L(N) \stackrel{\text{def}}{=} \sum_{\{n_i,\mu_i\}} \delta\big(N - \textstyle\sum_{i=1}^{L} n_i\big) \displaystyle\prod_{i=1}^{L} \pi^{\lambda}(n_i,\mu_i),$$

where δ denotes now the usual Dirac function. When ϕ is interpreted as a continuous variable, the properly normalized density-flow conditional probability distribution takes the form

$$P(\phi|d) = \frac{L}{1-d} \frac{Z_L\big[L\frac{d}{1-d}, L\frac{\phi}{1-d}\big]}{Z_L\big[L\frac{d}{1-d}\big]},$$

with

$$Z_L(N,\Phi) \stackrel{\text{def}}{=} \sum_{\{n_i,\mu_i\}} \delta\big(N - \textstyle\sum_{i=1}^{L} n_i\big)\, \delta\big(\Phi - \textstyle\sum_{i=1}^{L} \mu_i \mathbb{1}_{\{n_i>0\}}\big) \displaystyle\prod_{i=1}^{L} \pi^{\lambda}(n_i,\mu_i). \quad (14)$$

Note (by simple inspection, see e.g. [16]) that $P(\phi|d)$ is independent of λ. $Z_L(N)$ and $Z_L(N,\Phi)$ represent respectively the probability of having N vehicles and the joint probability for having at the same time N vehicles and a flow Φ, under the unconstrained product form. Under this product form, on general ground, we expect d and ϕ to satisfy a large deviation principle (see [30] for a recent review connecting large deviations to statistical physics), i.e. that there exist two rate functions $I(d)$ and $J(d,\phi)$ such that, for large L,

$$Z_L(N) \asymp e^{-LI(d)},$$

$$Z_L(N,\Phi) \asymp e^{-LJ(d,\phi)},$$

where "$\asymp$" stands as usual for logarithmic equivalence. This would lead to a large deviation version of the fundamental diagram of the form

$$P(\phi|d) \asymp e^{-LK(\phi|d)}, \quad \text{with} \quad K(\phi|d) \stackrel{\text{def}}{=} J(d,\phi) - I(d). \quad (15)$$

When there is one single constraint, like for $Z_L(N)$, the large deviation expression can be obtained by saddle point techniques [7,9]. For more than one constraint, it seems easier to work variationally. Consider the number $L_{n,\mu}$ of queues having n clients and service rate μ. Assuming μ runs over discrete values, the partition function (14) can be recast as:

$$\begin{aligned} Z_L(N,\Phi) = &\sum_{\{L_{n,\mu}\}} \frac{L!}{\prod_{n,\mu} L_{n,\mu}!} \prod_{n,\mu} \pi^{\lambda}(n,\mu)^{L_{n,\mu}} \\ &\times \delta\big(L - \sum_{n,\mu} L_{n,\mu}\big)\, \delta\big(N - \sum_{n,\mu} n L_{n,\mu}\big)\, \delta\big(\Phi - \sum_{n>0,\mu} \mu L_{n,\mu}\big) \end{aligned} \quad (16)$$

As can be seen explicitly in the following, when there is no condensation, i.e. no heavy tail in the $L(n,\mu)$ distribution, we are in the condition of the Sanov theorem [30] to get an asymptotic expression of the partition function. Namely,

$$Z_L(N,\Phi) \asymp \sum_{\{y(n,\mu)\}} \exp\Big(-L\mathscr{F}[y]\Big)\, \delta\big(C_1[y]\big)\, \delta\big(C_2[y]\big)\, \delta\big(C_3[y]\big), \quad (17)$$

with $y(n,\mu) \stackrel{\text{def}}{=} L_{n,\mu}/L$ the fraction of queues having n clients and service rate μ, $\mathscr{F}$ the large deviation functional and $C_{1,2,3}[y]$ the constraints:

$$C_1[y] = \sum_{n,\mu} y(n,\mu) - 1,$$

$$C_2[y] = \sum_{n,\mu} n y(n,\mu) - \frac{d}{1-d},$$

$$C_3[y] = \sum_{n>0,\mu} \mu y(n,\mu) - \frac{\phi}{1-d}.$$

This is obtained explicitly by making use of the Stirling formula in (16), keeping only the leading terms to get:

$$\mathscr{F}[y] = \sum_{n,\mu} y(n,\mu) \log \frac{y(n,\mu)}{\pi^\lambda(n,\mu)}. \tag{18}$$

The rate functions are then obtained using the contraction principle [30]:

$$I(d) = \inf_{\{y:C_1[y]=0,C_2[y]=0\}} \mathscr{F}[y],$$

$$J(d,\phi) = \inf_{\{y:C_1[y]=0,C_2[y]=0,C_3[y]=0\}} \mathscr{F}[y],$$

which applies here simply because $\mathscr{F}$ is convex and the constraints are linear. The variational principle consists in to looking for a distribution y minimizing the mutual information with the product form while satisfying the constraints; in this sense, it is equivalent to the standard mean field approximation. A simple generalization would then consist in, when the product form is not valid, to consider the 2-servers problem to build a joint-law based on pairs of nearest neighbor dependencies $\pi^\lambda(n_1,\mu_1;n_2,\mu_2)$, and to use the Bethe approximation corresponding to the following ansatz on the ring geometry ($(n_{L+1},\mu_{L+1}) \equiv (n_1,\mu_1)$):

$$P(\{n_i,\mu_i\}) = \frac{\delta\left(N - \sum_{i=1}^L n_i\right)}{Z_L(N)} \prod_{i=1}^L \pi^\lambda(n_{i+1},\mu_{i+1}|n_i,\mu_i),$$

The large deviation analysis would actually follow the same lines.

Let us introduce the moment and cumulant generating function g and h associated to π^λ,

$$g(s,t) \stackrel{\text{def}}{=} \sum_{n=0,\mu}^{\infty} \pi^\lambda(n,\mu) e^{sn+t\mu} \qquad \text{and} \qquad h(s,t) \stackrel{\text{def}}{=} \log[g(s,t)],$$

where it is assumed by convention that the rate μ is zero in absence of client. The stationary point then reads

$$y(n,\mu) = \frac{\pi^\lambda(n,\mu) e^{n\lambda_n + \mu\lambda_\mu}}{g(\lambda_n,\lambda_\mu)}, \tag{19}$$

where the Lagrange multipliers λ_n and λ_μ, associated to constraints C_2 and C_3, are implicitly given by

$$\begin{cases} \dfrac{\partial h}{\partial s}\big(\lambda_n(d,\phi),\lambda_\mu(d,\phi)\big) = \dfrac{d}{1-d}, \\ \dfrac{\partial h}{\partial t}\big(\lambda_n(d,\phi),\lambda_\mu(d,\phi)\big) = \dfrac{\phi}{1-d}. \end{cases} \tag{20}$$

Inserting the variational solution (19) into $\mathscr{F}$ leads to the resulting expression of the rate function as the Legendre transform of h

$$J(d,\phi) = \frac{d}{1-d}\lambda_n(d,\phi) + \frac{\phi}{1-d}\lambda_\mu(d,\phi) - h\big(\lambda_n(d,\phi),\lambda_\mu(d,\phi)\big),$$

which could actually be obtained directly from Cramer's theorem because $\{(n_i,\mu_i)\}$ are i.i.d. [30]. To obtain the other rate function $I(d)$, $\lambda_n'(d)$ is still associated to the first constraint, while λ_μ is set to zero, and we get

$$I(d) = \frac{d}{1-d}\lambda_n'(d) - h\big(\lambda_n'(d),0\big),$$

in terms of the cumulant generating function h, which completes the determination of the large deviation FD (15). We expect that analytical changes in the expression of $K(\phi|d)$ as d is varied could be associated to the onset of congestion or changes in the type of congestion identified in [17].

4.2 Gaussian fluctuations around the fundamental diagram

The deterministic part of the fundamental diagram is then obtained for $\phi(d)$ such that,

$$\frac{\partial J(d,\phi)}{\partial \phi}\Big|_{d,\phi(d)} = \frac{\lambda_\mu(d,\phi(d))}{1-d} = 0. \tag{21}$$

When considering small fluctuations rather than large deviations, we recover a central limit theorem version of the fundamental diagram. In that case, replacing in (17) $\mathscr{F}$ with its quadratic approximation around some stationary point value y^* of y yields the following expression for the partitions functions:

$$Z_L(N) \simeq \frac{\exp\big(-LI(d)\big)}{\sqrt{2\pi L h_{ss}(\lambda'_n,0)}}, \tag{22}$$

$$Z_L(N,\Phi) \simeq \frac{\exp\big(-LJ(d,\phi)\big)}{\sqrt{(2\pi L)^2 \det\big(H^\star(\lambda_n,\lambda_\mu)\big)}}, \tag{23}$$

with (λ_n,λ_μ) and λ'_n solving the constraints (20) for given values of d and ϕ, and the additional notation

$$H^\star(s,t) \overset{\text{def}}{=} \begin{bmatrix} h_{ss} & h_{st} \\ h_{ts} & h_{tt} \end{bmatrix}. \tag{24}$$

$H^\star(s,t)$ represents the covariance matrix between the charges of the queues and the flows and the subscripts to h are shorthand notations for the derivatives w.r.t. s and t. This matrix is actually the Hessian in the dual representation, associated to the Legendre transform of the free energy (see Appendix A for details). We obtain for the large deviation FD the expression

$$P(\phi|d) \asymp \sqrt{\frac{L}{2\pi(1-d)^2}\frac{h_{ss}(\lambda_n',0)}{\det\big(H^\star(\lambda_n,\lambda_\mu)\big)}}\exp\big(-LK(\phi|d)\big).$$

The Taylor expansion at second order in $\phi-\phi(d)$ at fixed density is performed using (21), so that

$$J\big(d,\phi(d)\big) = I(d),$$

and

$$\frac{\partial^2 J(d,\phi)}{\partial\phi^2}|_{d,\phi(d)} = \frac{1}{1-d}\frac{\partial\lambda_\mu(d,\phi)}{\partial\phi}|_{d,\phi(d)} = \frac{1}{(1-d)^2}\frac{h_{ss}(\lambda_n,\lambda_\mu)}{\det(H^\star(\lambda_n,\lambda_\mu))}.$$

The last equality is obtained by remarking that from the constraints (20) we have

$$\frac{\partial\lambda_n}{\partial\phi}h_{ss} + \frac{\partial\lambda_\mu}{\partial\phi}h_{st} = 0,$$

$$\frac{\partial\lambda_n}{\partial\phi}h_{st} + \frac{\partial\lambda_\mu}{\partial\phi}h_{tt} = \frac{1}{1-d},$$

from which $\partial\lambda_\mu/\partial\phi$ can be eliminated to yield the result. Therefore

$$K(\phi|d) = \frac{\big(\phi-\phi(d)\big)^2}{2(1-d)^2}\frac{h_{ss}}{\det(H^\star(\lambda_n,\lambda_\mu))} + o\big(\phi-\phi(d)\big)^2,$$

giving the Gaussian fluctuations of the FD. The variance as a function of the density then reads

$$\mathrm{Var}(\phi|d) = \frac{(1-d)^2}{L}\frac{\det(H^\star(\lambda_n(d),0))}{h_{ss}(\lambda_n(d),0)} = \frac{(1-d)^2}{L}\big({H^\star}^{-1}{}_{tt}\big)^{-1}. \qquad (25)$$

4.3 A special case: the TASEP

Let us apply this formula first to the simple M/M/1 server corresponding to one single speed level ($\mu_a=\mu_b=\mu$), i.e. a TASEP process. The rate of arrival λ is set by convenience to $\lambda = d\mu$. The cumulant-generating function then reads

$$h(s,t) = \log(1-d) + \log\frac{1+de^{-s}(e^{-tv}-1)}{1-de^{-s}}.$$

Therefore we have

$$\det(H^\star(0,0)) = \begin{vmatrix} \dfrac{d}{(1-d)^2} & d\mu \\ d\mu & d(1-d)\mu^2 \end{vmatrix} = \frac{d^3}{1-d}\mu^2,$$

while $\mathbb{E}[\phi|d] = \mu d(1-d)$. Equation (25) yields

$$\mathrm{Var}_{\mathrm{var}}(\phi|d) = \frac{1-d}{L} d^2(1-d)^2\mu^2. \tag{26}$$

By comparison, the direct result in the grand canonical ensemble, with expectation constraint on the density, is given by

$$\mathrm{Var}_{\mathrm{GC}}(\phi|d) = \frac{1-d}{L} h_{tt} = \frac{1-d}{L} d(1-d).$$

On the ring geometry, the computation for TASEP can be done directly as

$$\begin{aligned}\frac{1}{\mu^2}\mathrm{Var}(\phi|d) &= \mathrm{Var}\Big(\frac{1}{N+L}\sum_{i=1}^{N+L}\tau_i\bar{\tau}_{i+1}\Big)\\ &= \frac{(L+N)(L+N-3)}{(L+N)^2}\mathbb{E}(\tau_i\bar{\tau}_{i+1}\tau_j\bar{\tau}_{j+1})\\ &+ \frac{1}{L+N}\mathbb{E}(\tau_i\bar{\tau}_{i+1}) - \frac{1}{(L+N)^2}\Big(\mathbb{E}(\tau_i\bar{\tau}_{i+1})\Big)^2,\end{aligned}$$

with $\bar{\tau}_i = 1-\tau_i \in \{0,1\}$, $\tau_{N+L+1} = \tau_1$ and $\forall(i,j)$ s.t. $|i-j| \geq 2$. The expressions above are computed to give

$$\mathbb{E}(\tau_i\bar{\tau}_{i+1}) = \frac{NL}{(L+N)(L+N-1)} = d(1-d)(1+\frac{1}{L+N}) + o\big(\frac{1}{L+N}\big)$$

and

$$\begin{aligned}\mathbb{E}(\tau_i\bar{\tau}_{i+1}\tau_j\bar{\tau}_{j+1}) &= \frac{N(N-1)L(L-1)}{(L+N)(L+N-1)(L+N-2)(L+N-3)}\\ &= d^2(1-d)^2(1+\frac{6}{L+N}) - \frac{1}{L+N}d(1-d) + o\big(\frac{1}{L+N}\big),\end{aligned}$$

which coincides with (26) as expected.

5 Servers with two-state service rates for AB-TASEP

Let us now try to analyze the AB-TASEP model (1)-(4) using the tools developed in the preceding sections. Among the possible mappings discussed in Section 2.2, taking empty sites as servers should give more information on the jam structure than the mapping of type (i), because the long range correlations of the model are presumably associated to cluster formations. While this mapping is not exact here, let us try nevertheless to define a queueing process able to capture these correlations at least qualitatively.

5.1 Speed profile inside a single cluster

The first step that we take in this direction is to understand the dynamics and steady state regime of a single cluster of vehicles of size n_t, assuming that vehicles of type A [resp. B] join the queue with rate λ_a [resp. λ_b], while they leave the queue with rate μ_a [resp. μ_b]. Let $\lambda \stackrel{\text{def}}{=} \lambda_a + \lambda_b$ represents the intensity of the incoming process. In the bulk of the queue, according to the transition rules (3) and (4), only deceleration is possible, with rate δ. To study the speed profile, i.e. the speed labels of particles depending on their position inside a cluster, let us encode a given cluster configuration as a binary sequence $\{A_1,\ldots,A_n\}$, with $A_i = 1 - B_i, \in \{0,1\}$, corresponding to either a slow or a fast vehicle. n represents the size of the queue, the rightmost vehicle i.e. at the forefront of the jam has index n while index 0 corresponds to the leftmost vehicle (the last entered one). We remark first that the forefront of the jam is a moving interface, which next n moves are completely conditioned by the present state of the sequence. Let us forget about this interface for the moment and consider instead the distribution of the infinite sequence $\{A_1,A_2,\ldots\}$. Each time a new particle joins the cluster, the whole sequence is shifted by one unit ($A_i \longrightarrow A_{i+1}$), which induces dependencies between neighbors. Nevertheless, the structure of dependency is quite simple as, for example, single site marginals can be determined in close form. Letting $p_i(t) = P_t(A_i = 1)$, we have indeed

$$\frac{dp_i}{dt} = \lambda(p_{i-1} - p_i) - \delta p_i, \qquad \forall i > 1$$

$$\frac{dp_1}{dt} = \lambda_a(1 - p_1) - (\lambda_b + \delta)p_1,$$

with $\lambda \stackrel{\text{def}}{=} \lambda_a + \lambda_b$, and thus

$$p_1(t) = \frac{\lambda_a + p_1(0)e^{-(\lambda+\delta)t}}{\lambda + \delta}.$$

The generating function

$$\phi_t(z) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} p_{i+1}(t)z^i,$$

satisfies

$$\frac{d\phi_t(z)}{dt} = \big(\lambda z - (\lambda + \delta)\big)\phi_t(z) + \lambda_a,$$

with solution

$$\phi_t(z) = \frac{\lambda_a + \phi_0(z)e^{-(\lambda+\delta-\lambda z)t}}{\lambda + \delta - \lambda z}.$$

Correlations between nearest neighbors, $q_i(t) \stackrel{\text{def}}{=} P_t(A_i = 1, A_{i+1} = 1)$, satisfy a similar closed form equation:

$$\frac{dq_i}{dt} = \lambda(q_{i-1} - q_i) - 2\delta q_i, \qquad \forall i \geq 1$$

$$\frac{dq_1}{dt} = \lambda_a p_1(t) - (\lambda + 2\delta)q_1,$$

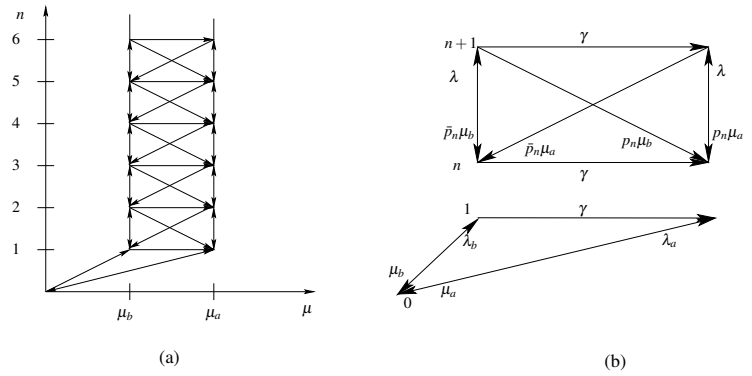


Fig. 6 (a) State flow diagram of a queue with stochastic service rate corresponding to AB-TASEP. (b) Details of the transition rates.

The generating function

$$\psi_t(z) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} q_i(t) z^i,$$

satisfies

$$\frac{d\psi_t(z)}{dt} = \big(\lambda z - (\lambda + 2\delta)\big)\psi_t(z) + \lambda_a p_1(t),$$

and is therefore equal to:

$$\psi_t(z) = \frac{\lambda_a^2}{(\lambda+\delta)(\lambda+2\delta-\lambda z)} + \Big[\psi_0(z) - \frac{\lambda_a^2}{(\lambda+\delta)(\lambda+2\delta-\lambda z)} + \frac{\lambda^2 p_1(0)(1-e^{-\lambda z t})}{(\lambda+\delta)\lambda z}\Big] e^{-(\lambda+2\delta-\lambda z)t}.$$

At steady state, we see that

$$p_i^{\infty} = \frac{\lambda_a}{\lambda}\Big(\frac{\lambda}{\lambda+\delta}\Big)^i \qquad \text{and} \qquad q_i^{\infty} = \frac{\lambda_a^2}{\lambda(\lambda+\delta)}\Big(\frac{\lambda}{\lambda+2\delta}\Big)^i,$$

indicating a significant level of correlations when δ/λ is $O(1)$.

5.2 Joint effective process $\big(n(t), \mu(t)\big)$

Since the front end interface of the cluster has no causal effect on the rest of the queue, except for the front vehicle which may accelerate with rate γ, we can consider the dynamics of the sequence independently of the motion of the front interface. Making the additional assumption of independence of the local speed labels in the bulk is rather crude regarding the results of the previous section, but nevertheless we look for a qualitative comparison. Let us write a master equation of the joint process $(n(t), \mu(t))$, i.e. the equation governing the evolution of

$P_t(n,\tau) = P(n(t) = n, \mu = \mu_a\tau + \mu_b\bar{\tau})$, the joint probability that the queue has n clients and its front car is of type A ($\tau = 1$) or B ($\bar{\tau} \stackrel{\text{def}}{=} 1 - \tau = 1$). Given

$$p_n \stackrel{\text{def}}{=} \frac{\lambda_a}{\lambda} r^n \qquad \text{with} \qquad r \stackrel{\text{def}}{=} \frac{\lambda}{\lambda + \delta}.$$

and $p_n(\tau) \stackrel{\text{def}}{=} p_n\tau + \bar{p}_n\bar{\tau}$, using $\bar{p}_n$ to denote $1 - p_n$, the master equation reads

$$\begin{aligned}
\frac{dP_t(n,\tau)}{dt} &= \lambda\big(P_t(n-1,\tau) - P_t(n,\tau)\big) + \big(\mu_a P_t(n+1,1) + \mu_b P_t(n+1,0)\big)p_n(\tau) \\
&\quad - (\mu_a\tau + \mu_b\bar{\tau})P_t(n,\tau) + \gamma(\tau - \bar{\tau})P_t(n,0), \qquad n \geq 2 \\
\frac{dP_t(1,\tau)}{dt} &= (\lambda_a\tau + \lambda_b\bar{\tau})P_t(0) - \lambda P_t(1,\tau) + \big(\mu_a P_t(2,1) + \mu_b P_t(2,0)\big)p_1(\tau) \\
&\quad - (\mu_a\tau + \mu_b\bar{\tau})P_t(1,\tau) + \gamma(\tau - \bar{\tau})P_t(1,0), \\
\frac{dP_t(0)}{dt} &= -\lambda P_t(0) + \mu_a P_t(1,1) + \mu_b P_t(1,0).
\end{aligned}$$

It is a special case of a queueing process with a 2-level dynamically coupled stochastic service rate, as defined in Section 3, which state-graph is represented on Figure 6. In the stationary regime, we denote

$$\pi_n^a \stackrel{\text{def}}{=} P(n(t) = n, \mu = \mu_a) \qquad \text{and} \qquad \pi_n^b \stackrel{\text{def}}{=} P(n(t) = n, \mu = \mu_b)$$

and $\pi_0 = P(n(t) = 0)$.

Proposition 1 *For $n > 1$, the following recursion holds*

$$\begin{bmatrix} \pi_{n+1}^a \\ \pi_{n+1}^b \end{bmatrix} = \frac{\lambda}{\mu_a\mu_b + \mu_a(\gamma + \lambda p_{n+1}) + \lambda\mu_b\bar{p}_{n+1}} \begin{bmatrix} \gamma + \mu_b + \lambda p_{n+1} & \gamma + \lambda p_{n+1} \\ \lambda\bar{p}_{n+1} & \mu_a + \lambda\bar{p}_{n+1} \end{bmatrix} \begin{bmatrix} \pi_n^a \\ \pi_n^b \end{bmatrix} \tag{27}$$

while for $n = 1$, we have

$$\pi_1^a = \frac{\lambda_a\mu_b + \lambda\gamma + \lambda^2 p_1}{\mu_a\mu_b + \lambda p_1\mu_a + \lambda\bar{p}_1\mu_b + \gamma\mu_a}\pi_0,$$

$$\pi_1^b = \frac{\lambda_b\mu_a + \lambda^2\bar{p}_1}{\mu_a\mu_b + \lambda p_1\mu_a + \lambda\bar{p}_1\mu_b + \gamma\mu_a}\pi_0.$$

Proof See Appendix B

Note that, for large n, because of a finite acceleration rate γ, the effective service rate μ^∞ has a limit slightly above μ_b. It is obtained as

$$\mu^\infty = \mu_b + \frac{\eta}{1+\eta}(\mu_a - \mu_b),$$

where η is the limit ratio

$$\eta = \lim_{n\to\infty} \frac{\pi_n^a}{\pi_n^b} = \frac{1}{2\lambda}\big(\sqrt{\Delta} + \gamma - \lambda + \mu_b - \mu_a\big) \quad \text{with} \quad \Delta \stackrel{\text{def}}{=} (\lambda - \gamma + \mu_a - \mu_b)^2 + 4\lambda\gamma$$

obtained from (27). The flow balance equation

$$\lambda = \sum_{n=1}^{\infty} \pi_n^a \mu_a + \pi_n^b \mu_b$$

is automatically fulfilled by virtue of the partial balance equation (30). Consider the generating functions

$$g_{a,b}(z) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} \pi_n^{a,b} z^n \qquad \text{and} \qquad g(z) \stackrel{\text{def}}{=} \pi_0 + g_a(z) + g_b(z).$$

Proposition 2 *(i) $g(z)$ satisfies the functional equation*

$$ug(rz) = -(z-z^+)(z-z^-)g(z) - vz + w$$

with $z^{\pm}$ given by

$$\begin{aligned} z^{\pm} &= \frac{1}{2\lambda}\left(\mu_a + \mu_b + \lambda + \gamma \pm \sqrt{\Delta}\right), \\ \lambda^2 u &= \lambda_a(\mu_a - \mu_b), \\ \lambda^2 v &= (\lambda_a \mu_a + \lambda_b \mu_b)\pi_0, \\ \lambda^2 w &= (\mu_a \mu_b + \lambda_a \mu_a + \lambda_b \mu_b + \gamma \mu_a)\pi_0. \end{aligned}$$

(ii) The solution reads:

$$g(z) = \sum_{n=0}^{\infty} (-u)^n \frac{w - v r^n z}{\prod_{k=0}^{n}(z r^k - z^+)(z r^k - z^-)}. \tag{28}$$

Proof See Appendix B for details.

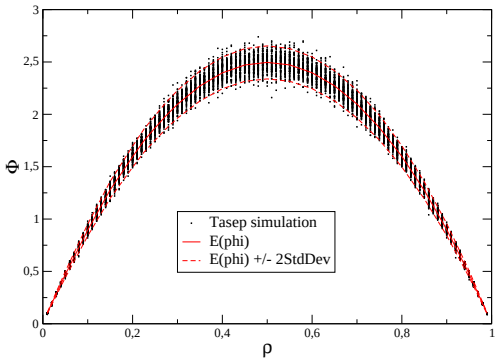
The reduced generating functions $g_{a,b}$ are then obtained using (33) and (34):

$$g_a(z) = \frac{\mu_b}{\mu_a - \mu_b}\pi_0 + \frac{\lambda z - \mu_b}{\mu_a - \mu_b} g(z) \quad \text{and} \quad g_b(z) = \frac{\mu_a}{\mu_b - \mu_a}\pi_0 + \frac{\lambda z - \mu_a}{\mu_b - \mu_a} g(z). \tag{29}$$

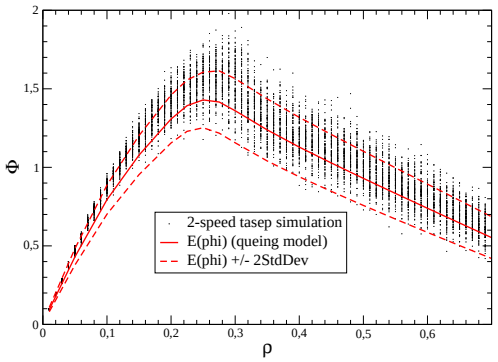
From these, it is then straightforward to obtain the $\pi_n^{a,b}$ upon using Cauchy integrals, as a sums of geometric laws, with parameters $r^k/z^{\pm}, k \in \mathbb{N}$. From the radius of convergence z^- of g, the limit of ergodicity is actually obtained for $z^- \geq 1$, i.e. for

$$\lambda \leq \mu_b + \gamma \frac{\mu_a - \mu_b}{\mu_a + \gamma},$$

which corresponds to μ_∞ encountered already in Section 2.3.



(a)

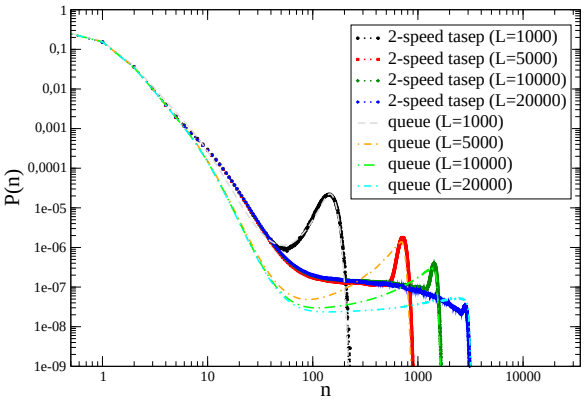


(b)

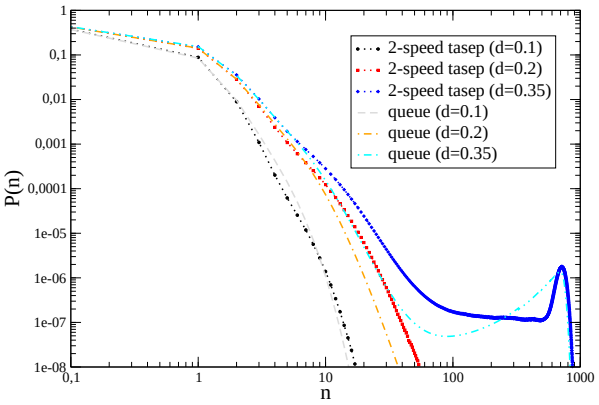
Fig. 7 (Color online) Comparison between the fundamental diagram simulated with exclusion processes and the one obtained from the corresponding approximate queueing process. The size is $L = 1000$, δ is set to zero on panel (a) to have a TASEP, while the parameters are set to $\mu_a = 10 \times \mu_b = 10 \times \gamma = 100 \times \delta$ on panel (b) for AB-TASEP.

5.3 Mean field estimates

If we consider then the closed tandem formed out of these effective queueing processes, the first observation is that the sufficient conditions given in Section 3.2 may not be fulfilled. As such, the large deviation estimation of the FD is going again to be only approximate, which comes in addition to the simplifying assumption made on the internal cluster structure.



(a)



(b)

Fig. 8 (Color online) Comparison of particle cluster vs queue's size distribution for set of parameters $\mu_a = 10\mu_b = 100$, $\gamma = 10$ and $\delta = 1$ for various sizes L with fixed density $d = 0.35$ (a) and various densities with fixed size $L = 5000$ (b), the number of servers being $(1-d)L$.

From (29), we write down the cumulant generating function

$$h(s,t) = \log\Big(\pi_0\big(1 + \frac{\mu_b e^{-\mu_a t} - \mu_a e^{-\mu_b t}}{\mu_a - \mu_b}\big) + \frac{(\mu_a - \lambda e^{-s})e^{-\mu_b t} + (\lambda e^{-s} - \mu_b)e^{-\mu_a t}}{\mu_a - \mu_b}\, g(s)\Big),$$

and the Hessian

$$\det\big(H(s^\star,0)\big) =$$

$$\frac{1}{g^2}\begin{bmatrix} gg''-g'^2 & g\big(\lambda e^{-s^\star}g+(1-\lambda e^{-s^\star})g'\big) \\ g\big(\lambda e^{-s^\star}g+(1-\lambda e^{-s^\star})g'\big) & g\big(\mu_a\mu_b\pi_0+(\lambda(\mu_a+\mu_b)e^{-s^\star}-\lambda^2-\mu_a\mu_b)g\big)\end{bmatrix},$$

where $s^\star$ is the point which satisfies

$$g'(s^\star) = -\rho.$$

From this, we can compute the small (Gaussian) fluctuations of the FD. A comparison with a simulation of the corresponding AB-TASEP is displayed on Figure 7.

In addition to the fundamental diagram, under the canonical ensemble constraint, it is also interesting to determine the single queue distribution. This can be obtained from the partition function [7] in the large deviation framework as

$$p_{CE}(n,\mu) = \pi^\lambda(n,\mu)\frac{Z_{L-1}(N-n)}{Z_L(N)}$$
$$\simeq \pi^\lambda(n,\mu)\exp\Big[L\big(h(s(d-x),0)-h(s(d),0)-\frac{d-x}{1-d-x}s(d-x)+\frac{d}{1-d}s(d)\big)\Big],$$

with $x \stackrel{\text{def}}{=} n/(N+L)$ and the density constraint

$$\frac{\partial h}{\partial s}(s(d),0) = \frac{d}{1-d},$$

satisfied by $s(d)$. A comparison of this queueing formulation with the AB-TASEP is given on Figure 8. In this case, a self-consistency condition has to be imposed to determine the parameter λ_a,

$$\lambda_a = \sum_{n=1}^{\infty}\pi_n^a\mu_a,$$

which otherwise would be free. The correspondence between the cluster size distribution observed on the AB-TASEP with the single queue distribution obtained from the generalized queueing process is rather accurate. In particular, in both cases, a bump is observed in the distributions at the same location, for small size systems. It indicates that condensation is observed as a finite size phenomena. In the thermodynamic limit, macroscopic jams are absent. In this respect, it is different from the type of condensation analyzed in [7], which is obtained under some conditions on the service rate, as a large deviation principle but with different scaling (speed in the large deviation terminology) than L. Concerning the FD, a larger discrepancy is observed between the AB-TASEP process and its corresponding effective queueing process. The reason for this, which is not visible on the cluster distributions, is that the ratios of fast over slow particles for a given size of cluster does not coincide. We do not know however whether this is due to the simplifying assumption on internal cluster structure or the neglect of correlations between queues. These results indicate anyway that the steady state of the AB-TASEP might be well accounted for by a little bit more refined joint cluster measure.

6 Conclusion and Perspectives

In this paper, motivated by questions raised in the context of traffic modelling, we have proposed a simple extension to the definition of the TASEP model to take into account acceleration and braking, thereby offering the possibility to study the effect of asymmetry between the two mechanisms within a simple model. With two different speed levels a rich dynamics is already present. An effective mapping on a generalized ZRP, that we have introduced in this paper, gives a way to interpret at least qualitatively the spontaneous jamming phenomena occurring in this model for some choice of parameters. In addition, we develop a large deviation formalism to study the fundamental diagram associated to this family of queueing processes. The discrepancy between numerical simulation and estimation (Figures 7 and 8) may originate from the various approximations we make, concerning the detail structure of the jams and the correlations between queues which are neglected when using our large deviation estimation of the FD. Our formalism can be adapted to the situation where the servers are not independent, when the solution of a 2 generalized server problem is accessible.

In the family of models that we considered, we found that condensation phenomena are associated to finite size effects on the ring geometry, and at least numerically the presence of a large macroscopic jam at large scale seems doomed to decay exponentially with no bump in the probability distribution. This is consistent with the approximate mapping on ZRP that we propose. Nevertheless, the situation could change when the number of speed levels is increased, as shown in our previous work [10] on the subject, and we suspect that synchronized flow can take place in this limit of large number of speed levels.

The situation where vehicles may enter or leave the system at very low rates, so that the density may change adiabatically and allows one to observe hysteresis effect, also deserves more studies. In particular, it would be interesting to find a relevant measure associated to trajectories in the FD, based for example on Brownian windings models proposed in [8], to quantify the hysteresis level in various regions of the FD.

Finally, the models considered here are limited to single lane traffic and could be easily generalized to multi-lane, using coupled exclusion processes like in [6].

7 Acknowledgements

We thank Guy Fayolle for useful discussions. We also thank referees for fruitful comments and references. This work was supported by the French National Research Agency (ANR) grant No ANR-08-SYSC-017.

A Complement to Section 4.2: constrained partition function and the dual Hessian

In this appendix, we explain the role played by the dual Hessian in the large deviation expression of the constrained partition functions. Assume that we have n linear constraints, written as $Cy = V$, where C is a $n \times d$ matrix, each line C_i corresponding to constraint i and V is a n-dimensional vector. We want to estimate the constrained partition function $Z_L(V)$ given in (17)

by integrating y, considered as a d-dimensional vector, over $\mathbb{R}^d$. To obtain the small fluctuations, we approximate first $\mathscr{F}$ at second order with respect to some reference point y^*

$$\mathscr{F}[y] = \mathscr{F}[y^*] + (y-y^*)^T \cdot \bigtriangledown \mathscr{F}[y^*] + (y-y^*)^T \bigtriangledown^2 \mathscr{F}[y^*](y-y^*) + o\big(\|y-y^*\|^2\big),$$

where y^* matches the constraints with given Lagrange multipliers Λ

$$y^*(\Lambda) = \underset{y}{\operatorname{argmin}} \Big[\mathscr{F}[y] - \Lambda^T C y\Big].$$

Since, in absence of constraints, $Z_L(V)$, as it stands in (16), is normalized to 1, at this order of approximation the partition function reads

$$Z_L(V) = L^{d/2}\sqrt{\frac{\det H}{(2\pi)^d}} \int \mathrm{d}^d y \, \exp\Big[-L\big(\mathscr{F}[y^*] + (y-y^*)^T \cdot \bigtriangledown \mathscr{F}[y^*] + \frac{1}{2}(y-y^*)^T H (y-y^*)\big)\Big]$$

$$\times \prod_{i=1}^{n} \delta(V_i - \sum_{j=1}^{d} C_{ij} y_j),$$

with $H = \bigtriangledown^2 \mathscr{F}[y^*]$ the Hessian taken at $y = y^*$. The dual Hessian is associated to the dual free energy $\mathscr{F}^\star$

$$\mathscr{F}^\star[\Lambda] \overset{\text{def}}{=} \mathscr{F}[y^*(\Lambda)] - \Lambda^T C y.$$

Using relations associated to the stationarity of y^*, it reads

$$H^\star = C^T H^{-1} C.$$

Due to the specific form (18) of $\mathscr{F}$, the Hessian turns out to be the covariance matrix between the various quantities V_i, as given in (24). A convenient way to express the constraints is to write

$$\prod_{i=1}^{n} \delta(V_i - \sum_{j=1}^{d} C_{ij} y_j) = \lim_{\alpha \to 0} \frac{1}{(2\pi\alpha)^{n/2}} e^{-\frac{1}{2\alpha}\|V - Cy\|^2},$$

so that a Gaussian integration over y can be performed. Since $\|V - Cy\| = \|C(y^* - y)\|$, we simply get

$$Z_L(V) = e^{-L\mathscr{F}[y^*]} \lim_{\alpha \to 0} \sqrt{\frac{\det H}{(2\pi\alpha)^n \det H_\alpha}} e^{\frac{L}{2} V^T H_\alpha^{-1} V}$$

with

$$H_\alpha \overset{\text{def}}{=} H + \frac{C^T C}{L\alpha}, \qquad V \overset{\text{def}}{=} \bigtriangledown \mathscr{F}(y^*).$$

Let P be the orthonormal projection on the subspace spanned by the set of constraint vectors C_k and $\bar{P} \overset{\text{def}}{=} 1 - P$, such that

$$C = CP \qquad \text{and} \qquad P\,C^T = C^T.$$

We have

$$\begin{aligned}
\det(H_\alpha) &= \det(H) \det\big(1 + \frac{C^T C}{L\alpha} H^{-1}\big) \\
&= \det(H) \det\big(P(1 + \frac{C^T C}{L\alpha} H^{-1})P + P\frac{C^T C}{L\alpha} H^{-1} \bar{P} + \bar{P}\bar{P}\big) \\
&= \det(H) \det_P\big(1 + \frac{C^T C}{L\alpha} H^{-1}\big) \\
&= \det(H) \Big(\frac{1}{(L\alpha)^n} \det(H^\star) + o\big(\frac{1}{\alpha^n}\big)\Big)
\end{aligned}$$

where $\det_P$ corresponds to the block determinant associated to the subspace of constraints, and

$$H^\star = C^T H^{-1} C$$

is the dual Hessian defined in (24). Concerning V, since y^* is a stationary point at least w.r.t. fluctuations orthogonal to the constraints, we have

$$\bar{P} V = 0.$$

Since

$$\begin{aligned} H_\alpha^{-1} &= H^{-1}\big(1 + \frac{C^T C}{L\alpha} H^{-1}\big)^{-1} \\ &= H^{-1}\big(PL\alpha H (C^T C)^{-1} P - P\bar{P} + \bar{P}\bar{P}\big) + o(\alpha), \end{aligned}$$

as a result

$$\lim_{\alpha \to 0} V^T H_\alpha^{-1} V = 0.$$

Finally, the Gaussian estimate of the partition function reads

$$Z_L(V) = \frac{L^{n/2}}{\sqrt{(2\pi)^n \det(H^\star)}} e^{-L\mathscr{F}[y^*]},$$

which yields in particular the expressions (22,23) up to a factor L^n caused by a different convention in the constraint definition.

B Complement to Section 5: steady state of the queueing process

From the master equations, taken at steady-state, we get the recurrence for $n \geq 2$

$$\lambda\big(\pi_{n-1}^a - \pi_n^a\big) + \big(\mu_a \pi_{n+1}^a + \mu_b \pi_{n+1}^b\big) p_n - \mu_a \pi_n^a + \gamma \pi_n^b = 0,$$

$$\lambda\big(\pi_{n-1}^b - \pi_n^b\big) + \big(\mu_a \pi_{n+1}^a + \mu_b \pi_{n+1}^b\big) \bar{p}_n - \mu_b \pi_n^b - \gamma \pi_n^b = 0.$$

For $n = 1$, it reads

$$\lambda_a \pi_0 - \lambda \pi_1^a + \big(\mu_a \pi_2^a + \mu_b \pi_2^b\big) p_1 - \mu_a \pi_1^a + \gamma \pi_1^b = 0,$$

$$\lambda_b \pi_0 - \lambda \pi_1^b + \big(\mu_a \pi_2^a + \mu_b \pi_2^b\big) \bar{p}_1 - \mu_b \pi_1^b - \gamma \pi_1^b = 0.$$

The sum of the two equation gives that the quantity $\lambda \pi_n - \big(\mu_a \pi_{n+1}^a + \mu_b \pi_{n+1}^b\big)$ is a constant independent of n which has to vanish, leading to the partial balance relation

$$\lambda \pi_n = \mu_a \pi_{n+1}^a + \mu_b \pi_{n+1}^b, \qquad \forall n \geq 0, \tag{30}$$

and the recurrence relations

$$\big(\mu_a + \lambda \bar{p}_{n+1}\big) \pi_{n+1}^a - (\gamma + \lambda p_{n+1}) \pi_{n+1}^b = \lambda \pi_n^a \tag{31}$$

$$-\lambda \bar{p}_{n+1} \pi_{n+1}^a + (\gamma + \mu_b + \lambda p_{n+1}) \pi_{n+1}^b = \lambda \pi_n^b, \tag{32}$$

which can be solved and yield (27). Recurrence (31,32) can also be used to obtain the following relations among the generating functions,

$$\lambda_a g_a(rz) + \lambda_a g_b(rz) = (\mu_a + \lambda - \lambda z) g_a(z) - \gamma g_b(z) - \lambda_a \pi_0 z,$$

$$\lambda_a g_a(rz) + \lambda_a g_b(rz) = \lambda g_a(z) - (\gamma + \mu_b - \lambda z) g_b(z) + \lambda_b \pi_0 z.$$

It follows that

$$g_b(z) = \frac{\lambda z}{\mu_b - \lambda z}\pi_0 + \frac{\lambda z - \mu_a}{\mu_b - \lambda z} g_a(z),$$

also related to (30), which rewrites

$$g_a(z) = \frac{\mu_b}{\mu_a - \mu_b}\pi_0 + \frac{\lambda z - \mu_b}{\mu_a - \mu_b} g(z), \tag{33}$$

$$g_b(z) = \frac{\mu_a}{\mu_b - \mu_a}\pi_0 + \frac{\lambda z - \mu_a}{\mu_b - \mu_a} g(z). \tag{34}$$

Thus $g(z)$ satisfies the functional equation given in Proposition 2. The solution is constructed as follows. Consider the infinite product

$$G(z) \stackrel{\text{def}}{=} \prod_{n=0}^{\infty} \big(1 - r^n \frac{z}{z^+}\big)\big(1 - r^n \frac{z}{z^-}\big).$$

We have

$$G(rz) = \frac{z^- z^+}{(z - z^-)(z - z^+)} G(z), \tag{35}$$

so a solution of the form

$$g(z) = \frac{C(z)}{G(z)}$$

has to verify

$$C(z) + \frac{u}{z^+ z^-} C(rz) = \frac{w - vz}{(z - z^-)(z - z^+)} G(z).$$

Equivalently, for any $n \geq 0$, this can be rewritten as

$$\begin{aligned}\big(-\frac{u}{z^+ z^-}\big)^n C(r^n z) - \big(-\frac{u}{z^+ z^-}\big)^{n+1} C(r^{n+1} z) &= \big(-\frac{u}{z^+ z^-}\big)^n \frac{w - vzr^n}{(zr^n - z^-)(zr^n - z^+)} G(r^n z) \\ &= (-u)^n \frac{w - vzr^n}{\prod_{k=0}^{n}(zr^k - z^-)(zr^k - z^+)} G(z),\end{aligned}$$

after using relation (35). Taking the sum leads to the solution (28), owing to

$$\lim_{n\to\infty} \big(-\frac{u}{z^+ z^-}\big)^n C(r^n z) = 0,$$

which can be checked after the fact.

References

1. Appert, C., Santen, L.: Boundary induced phase transitions in driven lattice gases with metastable states. PRL **86**, 2498 (2001)
2. Barlović, R., Santen, L., Schadschneider, A., Schreckenberg, M.: Metastable states in cellular automata for traffic flow. Eur. Phys. J. **B5**, 793 (1998)
3. Blank, M.: Hysteresis phenomenon in deterministic traffic flows. Journal of Statistical Physics **120**, 627–658 (2005)
4. Cantini, L.: Algebraic Bethe Ansatz for the two species ASEP with different hopping rates. J. Phys. A: Math. Theor. **41**, 095,001 (2008)
5. Derrida, B., Evans, M.R., Hakim, V., Pasquier, V.: Exact solution for 1d asymmetric exclusion model using a matrix formulation. J. Phys. A: Math. Gen. **26**, 1493–1517 (1993)
6. Evans, M., Kafri, Y., Sugden, K., Tailleur, J.: Phase diagrams of two-lane driven diffusive systems. J. Stat. Mech.: Theory and Experiment **2011**(06), P06,009 (2011)

7. Evans, M.R., Majumdar, S.N., Zia, R.K.P.: Canonical analysis of condensation in factorized steady states. Journal of Statistical Physics **123**(2), 357–390 (2006)
8. Fayolle, G., Furtlehner, C.: Dynamical windings of random walks and exclusion models. J. Stat. Phys. **114**, 229–260 (2004)
9. Fayolle, G., Lasgouttes, J.M.: Asymptotics and scalings for large closed product-form networks via the Central Limit Theorem. Markov Proc. Rel. Fields **2**(2), 317–348 (1996)
10. Furtlehner, C., Lasgouttes, J.: A queueing theory approach for a multi-speed exclusion process. In: Traffic and Granular Flow ' 07, pp. 129–138 (2007)
11. Golinelli, G., Mallick, K.: The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics. Journal of Physics A: Mathematical and General **39**(41), 12,679 (2006)
12. Harris, C.: Queues with state-dependant stochastic service rate. Operation Research **15**, 117–130 (1967)
13. Kafri, Y., Levine, E., Mukamel, D., Schütz, G.M., Török, J.: Criterion for phase separation in one-dimensional driven systems. Phys. Rev. Lett. **89**, 035,702 (2002)
14. Karimipour, V.: A multi-species asep and its relation to traffic flow. Phys. Rev. E **59**, 205 (1999)
15. Kaupužs, J., Mahnke, R., Harris, R.J.: Zero-range model of traffic flow. Phys. Rev. E **72**, 056,125 (2005)
16. Kelly, F.P.: Reversibility and stochastic networks. John Wiley & Sons Ltd. (1979). Wiley Series in Probability and Mathematical Statistics
17. Kerner, B.: The Physics of Traffic. Springer Verlag (2005)
18. Kipnis, C., Landim, C.: Scaling limits of Interacting Particles Systems. Springer-Verlag (1999)
19. Liggett, T.M.: Interacting Particle Systems. Springer, Berlin (2005)
20. Nagel, K., Paczuski, M.: Emergent traffic jams. Phys. Rev. E **51**(4), 2909–2918 (1995)
21. Nagel, K., Schreckenberg, M.: A cellular automaton model for freeway traffic. J. Phys. I,2 pp. 2221–2229 (1992)
22. O'Loan, O.J., Evans, M.R., Cates, M.E.: Jamming transition in a homogeneous one-dimensional system: The bus route model. Phys. Rev. E **58**, 1404–1418 (1998)
23. Samsonov, M., Furtlehner, C., Lasgouttes, J.: Exactly solvable stochastic processes for traffic modelling. Tech. Rep. 7278, INRIA (2010)
24. Schönhof, M., Helbing, D.: Criticism of three-phase traffic theory. Transportation Research **43**, 784–797 (2009)
25. Schreckenberg, M., Schadschneider, A., Nagel, K., Ito, N.: Discrete stochastic models for traffic flow. Phys. Rev. E **51**, 2339 (1995)
26. Schutz, G.M., Harris, R.J.: Hydrodynamics of the zero-range process in the condensation regime. J. Stat. Phys. **127**, 419 (2007)
27. Spitzer, F.: Interaction of Markov processes. Adv. Math. **5**, 246 (1970)
28. Sugiyama, Y., et al.: Traffic jams without bottlenecks: experimental evidence for the physical mechanism of the formation of a jam. New Journal of Physics **10**, 1–7 (2008)
29. Tóth, B., Valkó, B.: Onsager relations and Eulerian hydrodynamic limit for systems with several conservation laws. J. Stat. Phys. **112**, 497–521 (2003)
30. Touchette, H.: The large deviation approach to statistical mechanics. Physics Reports **478**, 1–69 (2009)